May 1969

EFFICIENCY COMPARISONS FOR TWO SUBSET SELECTION PROCEDURES

by

Shih Pau Yen

Technical Report No. 119

University of Minnesota

Minneapolis, Minnesota

Submitted as a Thesis to the Faculty of the Graduate School of the University of Minnesota in partial fulfillment of the requirement for degree of Master of Arts.

ACKNOWLEDGEMENT

I wish to express my sincere thanks to my adviser, Professor Milton Sobel for suggesting the topic of present research and for his encouragement and advice throughout the course of this work.

TABLE OF CONTENTS

CHAPTER I

Introduction and Summary

1.1. Statement of the Problem.

An important class of problems is concerned with the selection and ranking of k populations. The selection and ranking may be defined in terms of a parameter of the population which may physically represent the mean, the variance or some quantity or may be a function of these quantities. Usually in the selection and ranking problems, populations with large (or small) values of the parameters are considered desirable and, accordingly, we define the population with the largest (or smallest) of the unknown values of the k parameters to be the best. In many situations the experimenter is interested in selecting a subset containing the best ones.

The goal considered here is to select, from k populations, a subset containing at least one of the t best populations for given t and k $(1 \leq t < k)$ on the basis of a common fixed sample size n from each of the k populations $\pi_1, \pi_2, \ldots, \pi_k$. A correct selection (CS) is defined as the selection of any subset which contains at least one population π_i whose parameter value θ_i is among the t largest parameter value. For k normal populations with unknown mean and common variance $\sigma^2 = 1$ two procedures, R_S by Sobel [3] and R_G by Gupta [2], have been proposed, these procedures all satisfy the same basic probability requirement, namely that the probability of a correct selection (PCS) is at least P^* whenever the t best populations are at a distance (defined below) of at least δ^* from the remaining $k - t$

population; here $\delta^* \geq 0$ and $P^* < 1$ are specified.

In the present work, the relative efficiency of R_S and R_G is investigated. With both procedures satisfying the basic probability requirement, we note that the size of the selected subset is a random variable which takes on values $1, 2, 3, \ldots, k-t$. Hence one criterion of the efficiency of any procedure is the expected size of the selected subset $E(S)$. In this paper, we show that for $t = 1$ and $\delta^* = 0$, the procedure R_S is asymptotically ($P^* \to 1$) better (in the sense of $E(S)$) than the procedure R_G in a special configuration and that the reverse is true in another special configuration; these configurations to be defined later. Hence for $t = 1$ and $\delta^* = 0$ neither of these procedures is uniformly better than the other.

1.2. Notation and Requirement.

Let X_{ij} $(j = 1, 2, \ldots, n)$ denote the random sample from π_i $(i = 1, 2, \ldots, k)$. Observations between and within populations are all (mutually) independent. The distribution function of X_{ij} is assumed to be normal with mean μ_i and common known variance $\sigma^2 = 1$. The larger μ values are regarded as better. Let $\mu_{[1]} \leq \mu_{[2]} \leq \cdots \leq \mu_{[k]}$ denote ordered values of μ_i and let the ordered μ vector be denoted by $\vec{\mu} = (\mu_{[1]}, \mu_{[2]}, \ldots, \mu_{[k]})$. No priori information is assumed concerning the correct pairing of the π_i and $\mu_{[j]}$ $(i, j = 1, 2, \ldots, k)$.

A given distance function $\delta(\mu_2, \mu_1)$ measures the distance between populations with parameters $\mu_1 \leq \mu_2$. Then we define $\delta_t = \delta(\mu_{[k-t+1]}, \mu_{[k-t]})$ as the distance between the t-best populations and the remaining $k - t$ worst population. Let $\delta^*(\geq 0)$ and $P^*(< 1)$ denote specified constants. Let Ω denote the set of all $\vec{\mu}$ and $\Omega(\delta^*)$ denote the subset of Ω in which $\delta_t \geq \delta^*$.

Our basic probability requirement for the procedures R_S and R_G is

$$(1.1) \qquad P(CS) \geq P^* \qquad \text{for all } \vec{\mu} \in \Omega(\delta^*).$$

1.3. <u>Statement of Procedure R_G and R_S (for $t = 1$).</u>

Let $\bar{X}_i = \sum_{j=1}^{n} X_{ij}/n$ be the sample mean based on n observations corresponding to the population $N(\mu_i, 1)$ $(i = 1, 2,\ldots, k)$. Both procedures are based on the observations only through the sample mean $\bar{X}_i$ $(i = 1, 2,\ldots, k)$. Let $\bar{X}_{[1]} \leq \bar{X}_{[2]} \leq \cdots \leq \bar{X}_{[k]}$ denote ordered $\bar{X}$ values.

<u>Procedure R_G</u>: Put all π_i in the selected subset if and only if

$$(1.2) \qquad \bar{X}_i \geq \bar{X}_{[k]} - a_G$$

the constant $a_G > 0$ is set so that (1.1) is satisfied.

<u>Procedure R_S</u>: Put all π_i in the selected subset if and only if

$$(1.3) \qquad \bar{X}_i > \bar{X}_{[k-s+1]} - a_s$$

where we take the largest integer $s(1 \leq s \leq k - t)$ and also determine $a_s \geq 0$ so that (1.1) holds.

CHAPTER II

Asymptotic Comparison of the Two Procedures

In this chapter the comparisons of these two procedure R_S and R_G, based on the expected size of the selected subset, will be made for $t = 1$ and $\delta^* = 0$ under some special configurations. In the last section of this chapter, we show the main result of our works: If $k > 2$ and $j > \sqrt{\frac{k(k-1)}{2}}$ then as $P^* \to 1$, the procedure R_S is better than the procedure R_G for configuration C_{k-1} in the sense of having a smaller expected subset size.

2.1. Notations.

The following notations will be adopted:

(2.1) $$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad -\infty < z < \infty$$

(2.2) $$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \qquad -\infty < z < \infty$$

(2.3) $$\lambda = \sqrt{n}\,\delta$$

(2.4) $$\overline{X}_{(j)} = \text{sample mean from the population } N(\mu_{[j]}, 1)$$

(2.5) $$A_G = a_G\sqrt{n}$$

(2.6) $$A_S = a_S\sqrt{n}$$

2.2. The Expected Size of the Selected Subset.

Under a special configuration we investigate the expected subset size for the procedures R_S and R_G.

Definition 2.1. Special configurations C_j $(j = 1, 2, \ldots, k-1)$ are defined to be the set of vectors $\vec{\mu}$ for which

$$(2.7) \qquad \mu_{[1]} = \mu_{[2]} = \cdots = \mu_{[k-j]} = \mu \text{ (say) and}$$

$$\mu_{[k-j+1]} = \mu_{[k-j+2]} = \cdots = \mu_{[k]} = \mu + \delta$$

when $\delta \geq 0$. For $j = 1$ this is sometimes referred to as the generalized least favorable configuration.

The expected size of the selected subset under the procedure R_S and R_G for the configuration C_j $(j = 1, 2, \ldots, k-1)$ is denoted by

$$(2.8) \qquad E_S(S|C_j) \text{ and } E_G(S|C_j) \quad (j = 1, 2, \ldots, k-1), \text{ respectively.}$$

<u>Lemma 2.1</u>. For Φ defined in (2.2)

$$(2.9) \qquad \inf_{\vec{\mu}} P\{CS|\vec{\mu}, R_G\} = \int \Phi^{k-1}(x + A_G)\, d\Phi(x)$$

$$(2.10) \qquad \inf_{\vec{\mu}} P\{CS|\vec{\mu}, R_S\} = 1 - \int \Phi^{k-1}(x - A_S)\, d\Phi(x)$$

<u>Proof of the lemma:</u> By (1.2)

$$(2.11) \qquad P\{CS|R_G\} = P\{\overline{X}_{(k)} \geq \overline{X}_{[k]} - a_G\}$$

$$= P\{\max_{i \neq k} \overline{X}_{(i)} \leq \overline{X}_{(k)} + a_G\}$$

$$= \int \prod_{i=1}^{k-1} \Phi[x + \sqrt{n}\,(\mu_{[k]} - \mu_{[i]}) + A_G] d\Phi(x)$$

where $\overline{X}_{(k)}$ is defined by (2.4), since $\mu_{[k]} \geq \mu_{[i]}$ $(i = 1, \ldots, k-1)$ and Φ is strictly increasing the infimum of the $P(CS)$ is achieved when $\mu_{[1]} = \mu_{[2]} = \cdots = \mu_{[k]}$. Hence

$$(2.12) \qquad \inf_{\vec{\mu}} P\{CS|R_G\} = \int \Phi^{k-1}(x + A_G)\, d\Phi(x).$$

This proves (2.9). Similarly, we can prove (2.10).

For both procedures R_G and R_S we satisfy the same basic probability requirement (1.1) by setting the infimum equal to P^* i.e., under R_G

$$(2.13) \qquad \int \Phi^{k-1}(x + A_G)\, d\Phi(x) = P^*$$

and under R_S

$$(2.14) \qquad 1 - \int \Phi^{k-1}(x - A_S)\, d\Phi(x) = P^*$$

Thus, both A_G and A_S are functions of P^* and k. For specified P^* and k, we can obtain A_G and A_S which in turn gives us the values a_G and a_s by (2.5) and (2.6).

Before we prove the following results, let us define

$$(2.15) \qquad p_i = P\{\overline{X}_{(i)} = \overline{X}_{[2]}\} \quad (i = 1, 2, \ldots, k)$$

Obviously, $\sum_{i=1}^{k} p_i = 1$.

For P^* close to 1 the procedure R_S takes $s = k - 1$ and we then put π_i in selected subset if and only if

$$(2.16) \qquad \overline{X}_i > \overline{X}_{[2]} - a_{k-1}.$$

The following lemma gives the expected size of the selected subset under the procedure R_G and R_S.

Lemma 2.2. For the special configuration C_j $(j = 1, 2, \ldots, k - 1)$

$$(2.17) \qquad E_G(S|C_j) = (k - j) \int \Phi^{k-j-1}(x + A_G)\, \Phi^{j}(x + A_G - \lambda)\, d\Phi(x) + j \int \Phi^{j-1}(x + A_G)\, \Phi^{k-j}(x + A_G + \lambda)\, d\Phi(x)$$

$$(2.18)\qquad E_S(S|C_j) = k - (k-j)\int \Phi^{j}(x - A_S + \lambda)\,\Phi^{k-j-1}(x - A_S)\,d\Phi(x)$$
$$- j\int \Phi^{j-1}(x - A_S)\,\Phi^{k-j}(x - A_S - \lambda)\,d\Phi(x).$$

Proof: Define a random variable Y_i such that

$$Y_i = \begin{cases} 1 & \text{if } \pi_i \text{ is in the selected subset } (i = 1,2,\ldots,k) \\ 0 & \text{otherwise.} \end{cases}$$

Let $S = Y_1 + Y_2 + \ldots + Y_k = \sum_{i=1}^{k} Y_i$. Then S is the size of the selected subset, the expected size of the selected subset is given by

$$(2.19)\qquad E(S) = E(\sum_{i=1}^{k} Y_i) = \sum_{i=1}^{k} E(Y_i) = \sum_{i=1}^{k} P\{\pi_i \in \text{selected subset}\}.$$

By (1.2) and (2.19) we have for any j ($j=1, 2,\ldots, k-1$)

$$(2.20)\qquad E_G(S|C_j) = \sum_{i=1}^{k} P\{\overline{X}_{(i)} \geq \overline{X}_{[k]} - a_G\}$$
$$= (k-j)\,P\{\overline{X}_{(1)} \geq \overline{X}_{[k]} - a_G\} + j\,P\{\overline{X}_{(k)} \geq \overline{X}_{[k]} + a_G\}$$
$$= (k-j)\,P\{\max_{i\neq 1} \overline{X}_{(i)} \leq \overline{X}_{(1)} + a_G\}$$
$$+ j\,P\{\max_{i\neq k} \overline{X}_{(i)} \leq \overline{X}_{(k)} + a_G\}$$
$$= (k-j)\int \Phi^{k-j-1}(x + A_G)\,\Phi^{j}(x + A_G - \lambda)\,d\Phi(x)$$
$$+ j\int \Phi^{j-1}(x + A_G)\,\Phi^{k-j}(x + A_G + \lambda)\,d\Phi(x)$$

This proves (2.17).

By (2.16) and (2.19) we have for any $j(j = 1, 2,\ldots, k-1)$

$$(2.21)\qquad E_S(S|C_j) = \sum_{i=1}^{k} P\{\overline{X}_{(i)} > \overline{X}_{[2]} - a_s\}$$
$$= (k-j)\,P\{\overline{X}_{(1)} > \overline{X}_{[2]} - a_s\} + j\,P\{\overline{X}_{(k)} > \overline{X}_{[2]} - a_s\}.$$

If we denote these last two probabilities by T and T', respectively, then

$$(2.22)\qquad T = P\{\bar{X}_{(1)} > \bar{X}_{[2]} - a_s\}$$

$$= p_1 + (k-j-1)\, P\{\bar{X}_{(1)} > \bar{X}_{(2)} - a_s,\ \bar{X}_{(2)} = \bar{X}_{[2]}\}$$

$$+ j\, P\{\bar{X}_{(1)} > \bar{X}_{(k)} - a_s,\ \bar{X}_{(k)} = \bar{X}_{[2]}\}$$

$$= p_1 + (k-j-1)\, [P\{\bar{X}_{(2)} - a_s < \bar{X}_{(1)} < \bar{X}_{(2)} < \min_{i\neq 1,2} \bar{X}_{(i)}\}$$

$$+ (k-j-2)\, P\{\bar{X}_{(3)} < \bar{X}_{(2)} < \min_{i\neq 2,3} \bar{X}_{(i)}\}$$

$$+ j\, P\{\bar{X}_{(k)} < \bar{X}_{(2)} < \min_{i\neq 2,k} \bar{X}_{(i)}\}]$$

$$+ j\, [P\{\bar{X}_{(k)} - a_s < \bar{X}_{(1)} < \bar{X}_{(k)} < \min_{i\neq 1,k} \bar{X}_{(i)}\}$$

$$+ (k-j-1)\, P\{\bar{X}_{(2)} < \bar{X}_{(k)} < \min_{i\neq 2,k} \bar{X}_{(i)}\}$$

$$+ (j-1)\, P\{\bar{X}_{(k-1)} < \bar{X}_{(k)} < \min_{i=k-1,k} \bar{X}_{(i)}\}]$$

$$= p_1 + (k-j-1)[P\{\bar{X}_{(1)} < \bar{X}_{(2)} < \min_{i\neq 1,2} \bar{X}_{(i)}\}$$

$$- P\{\bar{X}_{(1)} + a_s < \bar{X}_{(2)} < \min_{i\neq 1,2} \bar{X}_{(i)}\}$$

$$+ (k-j-2)\, P\{\bar{X}_{(3)} < \bar{X}_{(2)} < \min_{i\neq 2,3} \bar{X}\}$$

$$+ j\, P\{\bar{X}_{(k)} < \bar{X}_{(2)} < \min_{i\neq 2,k} \bar{X}_{(i)}\}]$$

$$+ j[P\{\bar{X}_{(1)} < \bar{X}_{(k)} < \min_{i\neq 1,k} \bar{X}_{(i)}\} - P\{\bar{X}_{(1)} + a_s < \bar{X}_{(k)} < \min_{i\neq 1,k} \bar{X}_{(i)}\}$$

$$+ (k-j-1)\, P\{\bar{X}_{(2)} < \bar{X}_{(k)} < \min_{i\neq 2,k} \bar{X}_{(i)}\}$$

$$+ (j-1)\, P\{\bar{X}_{(k-1)} < \bar{X}_{(k)} < \min_{i\neq k-1,k} \bar{X}_{(i)}\}]$$

using the notation above

$$(2.23)\qquad T = p_1 + (k-j-1)\,p_2 + j\,p_k$$

$$- (k-j-1)\int_{-\infty}^{\infty}\Phi(x-A_s)[1-\Phi(x)]^{(k-j-2)}[1-\Phi(x-\lambda)]^j\,d\Phi(x)$$

$$- j\int_{-\infty}^{\infty}\Phi(x-A_s+\lambda)[1-\Phi(x+\lambda)]^{(k-j-1)}[1-\Phi(x)]^j\,d\Phi(x)$$

$$= 1 - (k-j-1)\int_{-\infty}^{\infty}\Phi(x-A_s)[1-\Phi(x)]^{(k-j-2)}[1-\Phi(x-\lambda)]^j\,d\Phi(x)$$

$$- j\int_{-\infty}^{\infty}\Phi(x-A_s+\lambda)[1-\Phi(x+\lambda)]^{(k-j-1)}[1-\Phi(x)]^{j-1}\,d\Phi(x)$$

The third term can be integrated by parts and one of the new terms cancels with the second term above, giving the result

$$(2.24)\qquad T = 1 - \int_{-\infty}^{\infty}[1 - \Phi(x+A_s-\lambda)]^j[1 - \Phi(x+A_s)]^{k-j-1}\,d\Phi(x).$$

By a similar arguement,

$$(2.25)\qquad T' = p_k + (k-j)[P\{\bar{X}_{(k)}-a_s \le \bar{X}_{(k)} < \bar{X}_{(1)} < \min_{i\neq 1,k}\bar{X}_{(i)}\}$$

$$+ (j-1)\,P\{\bar{X}_{(k-1)} < \bar{X}_{(1)} < \min_{i\neq 1,k-1}\bar{X}_{(i)}\}$$

$$+ (k-j-1)\,P\{\bar{X}_{(2)} < \bar{X}_{(1)} < \min_{i\neq 1,2}\bar{X}_{(i)}\}]$$

$$+ (j-1)[P\{\bar{X}_{(k)} - a_s \le \bar{X}_{(k)} < \bar{X}_{(k-1)} < \min_{i\neq k-1,k}\bar{X}_{(i)}\}$$

$$+ (j-2)\,P\{\bar{X}_{(k-2)} < \bar{X}_{(k-1)} < \min_{i\neq k-2,k-1}\bar{X}_{(i)}\}$$

$$+ (k-j)\,P\{\bar{X}_{(1)} < \bar{X}_{(k-1)} < \min_{i\neq 1,k-1}\bar{X}_{(i)}\}]$$

$$= 1-(k-j)\int\Phi(x-A_s-\lambda)[1-\Phi(x)]^{k-j-1}[1-\Phi(x-\lambda)]^{j-1}\,d\Phi(x)$$

$$- (j-1)\int\Phi(x-A_s)[1-\Phi(x+\lambda)]^{k-j}[1-\Phi(x)]^{j-2}\,d\Phi(x)$$

Again integrating by parts the last term above, we have after cancellation

$$(2.24) \qquad T' = 1 - \int_{-\infty}^{\infty} [1 - \Phi(x + A_S)]^{j-1}[1 - \Phi(x + A_S + \lambda)]^{k-j} \, d\Phi(x)$$

Substituting (2.24) and (2.26) in (2.21), we finally obtain

$$\begin{aligned}
(2.27) \qquad E_S(S|C_j) &= (k-j)[1 - \int \Phi^{j}(x-A_S+\lambda)\, \Phi^{k-j-1}(x-A_S) \, d\Phi(x)] \\
&\quad + j[1 - \int \Phi^{k-j}(x-A_S-\lambda)\Phi^{j-1}(x-A_S) \, d\Phi(x)] \\
&= k - (k-j)\int \Phi^{j}(x-A_S+\lambda)\, \Phi^{k-j-1}(x-A_S) \, d\Phi(x) \\
&\quad - j \int \Phi^{k-j}(x-A_S-\lambda)\, \Phi^{j-1}(x-A_S) \, d\Phi(x)
\end{aligned}$$

This proves (2.18).

<u>Corollary 2.1</u>. For $\lambda = 0$ we have for the equal parameter (EP) configuration

$$(2.28) \qquad E_G(S|EP) = kP^*$$

$$(2.29) \qquad E_S(S|EP) = kP^*$$

and for $\lambda \to \infty$ we have for $j = 1, 2, \ldots, k - 1$

$$(2.30) \qquad \lim_{\lambda\to\infty} E_G(S|C_j) = j \int \Phi^{j-1}(x+A_G) \, d\Phi(x)$$

$$(2.31) \qquad \lim_{\lambda\to\infty} E_S(S|C_j) = k - (k-j) \int \Phi^{k-j-1}(x-A_S) \, d\Phi(x).$$

<u>Proof</u>. Substituting $\lambda = 0$ in (2.17) and (2.18) and using (2.13) and (2.14) we obtain

$$(2.32) \qquad E_G(S|EP) = k \int \Phi^{k-1}(x+A_G) \, d\Phi(x) = kP^*,$$

$$(2.33) \qquad E_S(S|EP) = k[1 - \int \Phi^{k-1}(x-A_S) \, d\Phi(x)] = kP^*.$$

Substituting $\lambda = \infty$ in (2.17) and (2.18) respectively, we obtain (2.30) and (2.31) as above; this proves the corollary. For $k = 5$ the value of (2.30) and (2.31) can be found in table IA, IB, IC, and ID for selected values of P^*.

2.3. An Asymptotic Comparison of the Two Procedures.

Our plan is to start with the expression in (2.13) and (2.14) and find an asymptotic ($P^* \to 1$) expression for A_G and A_S, and then substitute them into (2.17) and (2.18) to evaluate the expected subset size as $P^* \to 1$.

Lemma 2.3. As $P^* \to 1$

$$(2.34) \qquad A_G \sim 2\sqrt{\ln(\frac{1}{1-P^*})}$$

$$(2.35) \qquad A_S \sim \sqrt{\frac{2k}{k-1} \ln(\frac{1}{1-P^*})}$$

Proof. Fom (2.13) we obtain

$$(2.36) \qquad \int_{-\infty}^{\infty} \Phi^{k-1}(x + A_G)\, d\Phi(x) = P^*.$$

Using $\sim$ (or a.e.) for asymptotic equivalence and the fact that

$$(2.37) \qquad (1 - \epsilon)^k \sim 1 - k\,\epsilon \qquad \text{for } \epsilon \to 0 \text{ and fixed } k,$$

we rewrite (2.36) as

$$(2.38) \qquad P^* = \int_{-\infty}^{\infty} \{1-[1-\Phi(x+A_G)]\}^{k-1} d\Phi(x) \sim \int_{-\infty}^{\infty} \{1-(k-1)[1-\Phi(x+A_G)]\}\, d\Phi(x)$$
$$\sim 1-(k-1)\int_{-\infty}^{\infty} [1-\Phi(x+A_G)]\, d\Phi(x).$$

Hence we obtain after integration

$$(2.39) \qquad (k - 1)[1 - \Phi(\frac{A_G}{\sqrt{2}})] = 1 - P^*.$$

Using the first term of the Feller-Laplace expansion [1] for the "tail" of the normal distribution

$$(2.40) \qquad 1 - \Phi(x) \sim \frac{\varphi(x)}{x} \qquad \text{for} \quad x \to \infty$$

(2.39) becomes

$$(2.41) \qquad \frac{\varphi\left(\frac{A_G}{\sqrt{2}}\right)}{\frac{A_G}{\sqrt{2}}} \sim \frac{1 - P^*}{k - 1} .$$

Taking ln of both sides in (2.41) gives the a.e. form

$$(2.42) \qquad A_G \sim 2\sqrt{\ln\left(\frac{1}{1-P^*}\right)}.$$

The terms neglected do not affect our result, since the magnitude of these terms is much smaller than the terms kept for P^* close to 1.

For (2.14), we obtain

$$(2.43) \qquad \int \Phi^{k-1}(x - A_S)\, d\Phi(x) = 1 - P^*$$

Using the usual o-notation, we write the left side of (2.43) as

$$(2.44) \qquad \int_{-\theta A_S}^{\theta A_S} \Phi^{k-1}(x-A_S)\, d\Phi(x) + o(\exp\{-(\theta A_S)^2/2\})$$

where θ is such that $\frac{(k-1)}{k} < \theta^2 < 1$, using the first term of the Feller-Laplace expansion for the "tail" of the normal distribution, we write the integral in (2.44) without the constant factor, as

$$(2.45) \qquad \int_{-\theta A_S}^{\theta A_S} \frac{\varphi^{k-1}(x-A_S)}{|x-A_S|^{k-1}}\, d\Phi(x) \sim \frac{C}{A_S^{k-1}} \int_{-\theta A_S}^{\theta A_S} \varphi^{k-1}(x-A_S)\, d\Phi(x)$$

$$\sim \frac{C}{A_S^{k-1}} \int_{-\infty}^{\infty} \varphi^{k-1}(x-A_S)\, d\Phi(x) + o(\exp\{-(\theta A_S)^2/2\})$$

Hence, neglecting the terms of order at most $o(\exp\{-(\theta A_S)^2/2\})$ we obtain from (2.45) the a.e. form

$$(2.46) \qquad \int_{-\infty}^{\infty} \varphi^{k-1}(x - A_S)\, d\Phi(x) \sim C'(1-P^*)\, A_S^{k-1}.$$

A straight forward completion of the square and integrating in (3.46) gives the a.e. form

$$(2.47) \qquad C''A_S^{k-1}(1 - P^*) \sim \frac{e^{-\frac{(k-1)}{2k} A_S^2}}{(\sqrt{2\pi})^{k-1}}$$

From (3.47) taking ln on both sides, we easily obtain our final asymptotic $(P^* \to 1)$ expression for A_S

$$(2.48) \qquad A_S \sim \sqrt{\frac{2k}{k-1} \ln\left(\frac{1}{1-P^*}\right)} \ .$$

<u>Theorem 1</u>: For the configuration C_j as $P^* \to 1$

$$(2.49) \qquad k - E_S(S|C_j) > k - E_G(S|C_j)$$

when

$$(2.50) \qquad j > \sqrt{\frac{k(k-1)}{2}}$$

and the inequality in (2.49) is reversed if the inequality of (2.50) is reversed.

<u>Proof</u>. From (2.17)

$$(2.51) \qquad k-E_S(S|C_j) = (k-j)\int \Phi^{j}(x-A_S+\lambda)\, \Phi^{k-j-1}(x-A_S)\, d\Phi(x)$$
$$+ j \int \Phi^{j-1}(x-A_S)\Phi^{k-j}(x-A_S-\lambda)\, d\Phi(x)$$

Using the Feller-Laplace expansion for the "tail" of the normal c.d.f. in (2.51), dropping the denominator, then completing the square and neglecting the error term, $o(\exp\{-\theta^2(A - \lambda)^2/2\})$, we obtain

$$(2.52)\qquad k-E_S(S|C_j) \sim (k-j)\int \frac{\varphi(x-A_S+\lambda)^j}{|x-A_S+\lambda|^j}\,\frac{\varphi(x - A_S)^{k-j-i}}{|x-A_S+\lambda|^{k-j-1}}\,\varphi(x)dx$$

$$+ j\int \frac{\varphi(x-A_S)^{j-1}}{|x-A_S|^{j-1}}\,\frac{\varphi(x-A_S-\lambda)^{k-j}}{|x-A_S-\lambda|^{k-j}}\,\varphi(x)dx$$

$$\sim \frac{C}{A_S^{k-1}}\, e^{-\frac{1}{2}\frac{(k-1)}{k}A_S^2}\, e^{\frac{j\lambda A_S}{k}}$$

It is shown in lemma 2.3 that for $P^* \to 1$

$$(2.53)\qquad A_S \sim \sqrt{2\left(\frac{k}{k-1}\right)\ln\left(\frac{1}{1-P^*}\right)}$$

and applying this to (2.52) gives the final form for procedure R_S.

$$(2.54)\qquad k - E_S(S|C_j) \sim \frac{C}{A_S^{k-1}}\,(1 - P^*)\,\exp\{\lambda\sqrt{2\frac{j^2}{k(k-1)}\ln\left(\frac{1}{1-P^*}\right)}\}.$$

For procedure R_G it is shown in lemma 2.2 that

$$(2.55)\qquad E_G(S|C_j) = (k-j)\int \Phi(x+A_G)^{k-j-1}\,\Phi(x+A_G-\lambda)^j\,d\Phi(x)$$

$$+ j\int \Phi(x + A_G)^{j-1}\,\Phi(x + A_G + \lambda)^{k-j}\,d\Phi(x)$$

$$\sim (k-j)\int \{1-(k-j-1)[1-\Phi(x+A_G)]\}\{1-j[1-\Phi(x+A_G-\lambda)]\}\,d\Phi(x)$$

$$+ j\int \{1-(j-1)[1-\Phi(x+A_G)]\}\{1(k-j)[1-\Phi(x+A_G+\lambda)]\}\,d\Phi(x)$$

$$k - E_G(S|C_j) \sim C[1-\Phi(\frac{A_G}{\sqrt{2}})] + C'[1-\Phi(\frac{A_G-\lambda}{\sqrt{2}})] + C''[1-\Phi(\frac{A_G+\lambda}{\sqrt{2}})]$$

the terms which are neglected above do not affect our result, since the magnitude of the neglected terms are smaller than those kept as $P^* \to 1$.

Using the Feller-Laplace expansion for the "tail" of the normal distribution, the first term and third terms can be neglected for the same reason as above.

We obtain from (2.55)

$$(2.57) \qquad k - E_G(S|C_j) \sim \frac{C'}{A_G} \exp\{-\frac{A_G^2}{4} + \frac{A_G\lambda}{2}\}$$

It is shown in lemma 2.3 that for $P^* \to 1$

$$A_G \sim 2\sqrt{\ln(\frac{1}{1-P^*})}$$

and applying this to (2.57) gives the final form from procedure R_G

$$(2.58) \qquad k-E_G(S|C_j) \sim \frac{C'}{A_G}(1 - P^*) \exp\{\lambda\sqrt{\ln(\frac{1}{1-P^*})}\}.$$

It follows from (2.54) and (2.58) that (2.49) hold in theorem 1 when

$$(2.59) \qquad j > \sqrt{\frac{k(k-1)}{2}}$$

and the inequality in (2.49) is reversed when

$$(2.60) \qquad j < \sqrt{\frac{k(k-1)}{2}} .$$

For $k = 2$ the procedures are identical and (2.50) is vacuous. For $K = 3, 4, 5$ and 6 the following table describes for which configurations R_S is asymptotically better and for which R_G is asymptotically better.

Procedure	k value			
	3	4	5	6
R_S	C_2	C_3	C_4	C_4, C_5
R_G	C_1	C_1, C_2	C_1, C_2, C_3	C_1, C_2, C_3

Note that (2.50) always holds for C_{k-1} for all $k > 2$ and the reverse always holds for C_1 for all $k > 2$. The exact expected subset sizes for $K = 5$ under the configurations C_j $(j = 1, 2, 3, 4)$ are given in tables IA, IB, IC and ID respectively. In table ID, we note for that $P^* = 0.9990$ and 0.9999 the expected value of the subset size for procedure R_S is smaller than for procedure R_G in the special configuration C_4 for $\lambda = 1, 2, 3$.

TABLE IA

Comparison of $E\{S\}$ Values for the Two Procedures R_S and R_G Under the Special Configuration C_1.

Normal Location-Parameter Problem with Common $\sigma^2 = 1$

($k = 5$, $t = 1$, $\lambda = \delta\sqrt{n}$, $A_S = a_S\sqrt{n}$, $A_G = a_G\sqrt{n}$)

λ	Procedure	P^*			
		0.9000	0.9900	0.9990	0.9999
0	R_S	4.5000	4.9500	4.9950	4.9995
	R_G	4.5000	4.9500	4.9950	4.9995
1	R_S	4.4800	4.9409	4.9934	4.9992
	R_G	4.2869	4.8980	4.9858	4.9981
2	R_S	4.4639	4.9339	4.9921	4.9990
	R_G	3.5756	4.6370	4.9182	4.9830
3	R_S	4.4605	4.9324	4.9917	4.9990
	R_G	2.5329	3.9638	4.6431	4.8926
∞	R_S	4.4597	4.9322	4.9917	4.9990
	R_G	1	1	1	1

Formulas for this table are

i) $E(S, R_S) = 5 - 4 \int \Phi(x-A_S+\lambda)\, \Phi^3(X-A_S)\, d\Phi(x) - \int \Phi^4(x-\lambda-A_S)\, d\Phi(x)$

ii) $E(S, R_G) = 4 \int \Phi^3(x+A_G)\, \Phi(x+A_G-\lambda)\, d\Phi(x) + \int \Phi^4(x+A_S+\lambda)\, d\Phi(x)$

TABLE IB

Comparison of $E\{S\}$ Values for the Two Procedures R_S and R_G Under the Special Configuration C_2.

Normal Location-Parameter Problem with Common $\sigma^2 = 1$

$(k = 5,\quad t = 1,\quad \lambda = \delta\sqrt{n},\quad A_S = a_S\sqrt{n},\quad A_G = a_G\sqrt{n}\ )$

λ	Procedure	P^*			
		0.9000	0.9900	0.9990	0.9999
0	R_S	4.5000	4.9500	4.9950	4.9995
	R_G	4.5000	4.9500	4.9950	4.9995
1	R_S	4.4598	4.9297	4.9911	4.9989
	R_G	4.2188	4.8767	4.9817	4.9974
2	R_S	4.4145	4.9053	4.9856	4.9978
	R_G	3.4563	4.5468	4.8891	4.9759
3	R_S	4.4013	4.8969	4.9832	4.9726
	R_G	2.6192	3.8242	4.5477	4.8542
∞	R_S	4.3994	4.8956	4.9826	4.9971
	R_G	1.9339	1.9944	1.9995	2.000

Formulas for this table are

i) $E(S,\, R_S) \;=\; 5-3\int \Phi^2(x-A_S+\lambda)\, \Phi^2(x-A_S)\, d\Phi(x) - 2\int \Phi(x-A_S)\, \Phi^3(x-\lambda-A_S)\,d\Phi(x)$

ii) $E(S,\, R_G) \;=\; 3\int \Phi^2(x+A_G)\, \Phi^2(x+A_G-\lambda)\, d\Phi(x) + 2\int \Phi(x+A_G)\, \Phi^3(x+A_G+\lambda)\, d\Phi(x)$

TABLE IC

Comparison of $E\{S\}$ Values for Two Procedures R_S and R_G Under the Special Configuration C_3.

Normal Location-Parameter Problem with Common $\sigma^2 = 1$

$(k = 5,\quad t = 1,\quad \lambda = \delta\sqrt{n},\quad A_S = a_S\sqrt{n},\quad A_G = a_G\sqrt{n}\ )$

λ	Procedure	P^*			
		0.9000	0.9900	0.9990	0.9999
0	R_S	4.5000	4.9500	4.9950	4.9995
	R_G	4.5000	4.9500	4.9950	4.9995
1	R_S	4.4436	4.9178	4.9881	4.9983
	R_G	4.2491	4.8807	4.9820	4.9974
2	R_S	4.1639	4.8484	4.9684	4.9938
	R_G	3.6676	4.5947	4.8964	4.9769
3	R_S	4.3007	4.8038	4.9518	4.9879
	R_G	3.1398	4.0368	4.6036	4.8659
∞	R_S	4.2910	4.7896	4.9413	4.9837
	R_G	2.8222	2.9839	2.9984	3.0000

Formulas for this table are

i) $$E(S, R_S) = 5-2\int \Phi^3(x-A_S+\lambda)\,\Phi(x-A_S)\,d\Phi(x)-3\int \Phi^2(x-A_S)\,\Phi^2(x-A_S-\lambda)\,d\Phi(x)$$

ii) $$E(S, R_G) = 2\int \Phi(x+A_G)\,\Phi^3(x+A_G-\lambda)\,d\Phi(x) + 3\int \Phi^2(x+A_G)\,\Phi^2(x+A_G+\lambda)\,d\Phi(x)$$

TABLE ID

Comparison of $E\{S\}$ Values for the Two Procedures R_S and R_G Under the Special Configuration C_4.

Normal Location-Parameter Problem with Common $\sigma^2 = 1$

$(k = 5, \quad t = 1, \quad \lambda = \delta\sqrt{n}, \quad A_S = a_S\sqrt{n}, \quad A_G = a_G\sqrt{n} \;)$

λ	Procedure	P^*			
		0.9000	0.9900	0.9990	0.9999
0	R_S	4.5000	4.9500	4.9950	4.9995
	R_G	4.5000	4.9500	4.9950	4.9995
1	R_S	4.4443	4.9132	4.9857	4.9978
	R_G	4.3491	4.9062	4.9866	4.9981
2	R_S	4.2717	4.7380	4.9194	4.9780
	R_G	4.0400	4.7375	4.9334	4.9850
3	R_S	4.1009	4.4312	4.7160	4.8802
	R_G	3.7956	4.4364	4.7618	4.9166
∞	R_S	4	4	4	4
	R_G	3.6764	3.9688	3.9972	3.9996

Formulas for this table are

i) $$E\{S, R_S\} = 5 - \int \Phi^4(x-A_S+\lambda)\, d\Phi(x) - 4\int \Phi^3(x-A_S)\, \Phi(X-A_S-\lambda)\, d\Phi(x)$$

ii) $$E\{S, R_G\} = \int \Phi^4(x + A_G-\lambda)\, d\Phi(x) + 4\int \Phi^3(x+A_G)\, \Phi(X+A_G+\lambda)\, d\Phi(x)$$

BIBLIOGRAPHY

[1] Feller, W. (1961). An Introduction to Probability Theory and its Applications, Vol I (2nd Edition). John Wiley and Sons, Inc., New York.

[2] Gupta, S. S. (1965). On some multiple decision (selection and ranking) rules. Technometrics 7, 225-245.

[3] Sobel, M. (1969). Selecting a subset containing at least one of the t best populations. Multivariate Analysis Vol. II Proceedings of an International Symposium. Ed. by P. R. Krishnaiah. Academic Press, New York.